

Method for realizing polylink flow equalization

Publication number: CN1484413

Publication date: 2004-03-24

Inventor: HU CHUNSHENG (CN)

Applicant: SHENZHEN ZHONGXING COMM CO LTD (CN)

Classification:

- international: *H04B7/005; H04L12/08; H04L12/24; H04L12/56; H04B7/005; H04L12/08; H04L12/24; H04L12/56; (IPC1-7): H04L12/56; H04B7/005; H04L12/08; H04L12/24*

- European:

Application number: CN20021037079 20020918

Priority number(s): CN20021037079 20020918

Also published as:



CN100336363C

[Report a data error](#)

Abstract of CN1484413

This invention discloses a method for realizing balance of aggregate link flow including the following steps: receiving a single broadcast data packet, if there are matched items in the address list, then they are forwarded; if there is no matched address item, it is sent to an address study module which forecasts the future flow of each port according to its historical flow information to forward the said data packet via the port with the least flow in the future and renew the address list, broadly applied in an aggregate link bound by multiple physical links and a logic link between network devices, overcoming the shortcoming of unbalanced loading for different ports decided according to certain bits of a data packet address.

Data supplied from the *esp@cenet* database - Worldwide

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

H04L 12/56

H04L 12/24 H04L 12/08

H04B 7/005



[12] 发明专利申请公开说明书

[21] 申请号 02137079.6

[43] 公开日 2004 年 3 月 24 日

[11] 公开号 CN 1484413A

[22] 申请日 2002.9.18 [21] 申请号 02137079.6

[71] 申请人 深圳市中兴通讯股份有限公司

地址 518057 深圳市南山区高新技术产业园
科技南路中兴通讯大厦法律部

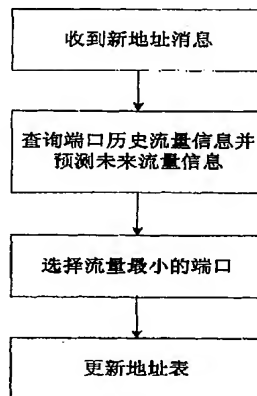
[72] 发明人 胡春生

权利要求书 1 页 说明书 4 页 附图 2 页

[54] 发明名称 一种实现聚合链路流量均衡的方法

[57] 摘要

本发明公开了一种实现聚合链路流量均衡的方法，包括以下处理步骤：接收一个单播数据包，如果地址表中有相匹配的表项则转发；如果地址表中没有相匹配的地址项，则送往地址学习模块；地址学习模块根据端口的历史流量信息预测各端口未来的流量；把该数据包通过未来流量最小的端口转发，并更新地址表。本发明可广泛应用于网络设备之间多条物理链路捆绑成一条逻辑链路的聚合链路中。采用本发明的实现聚合链路流量均衡的方法，可以克服以往根据数据包地址的某几位来决定转发端口可能使不同端口负载不均衡的缺点，使转发数据包时可以更均衡地在物理链路上分配流量，更充分地利用物理链路提供的带宽。



ISSN 1000-4274

1. 一种实现聚合链路流量均衡的方法，包括以下处理步骤：

- 步骤一，接收一个单播数据包，如果地址表中有相匹配的表项则转发；
 步骤二，如果地址表中没有相匹配的地址项，则送往地址学习模块；
 步骤三，地址学习模块根据端口的历史流量信息预测各端口未来的流量；
 步骤四，把该数据包通过未来流量最小的端口转发，并更新地址表。

2. 根据权利要求 1 所述的实现聚合链路流量均衡的方法，其特征在于，所述步骤三中预测端口未来流量具体包括以下处理步骤：

- a) 采用定时查询的方法，每隔 T 秒读取计数器中累计的端口发送字节数，
 假设端口 i 第 j 次查询的结果是 C_{ij} ；
 b) 计算端口 i 第 j 次查询前 T 秒的平均速率，计算公式为：

$$S_{i,j} = \frac{C_{i,j} - C_{i,j-1}}{T}$$

- c) 采用过去几个时间段流量的加权平均值来估计未来的流量。

3. 根据权利要求 2 所述的实现聚合链路流量均衡的方法，其特征在于，可以采用两个时间段的流量预测未来 T 秒的平均速率，预测公式为：

$$S'_{i,j} = \alpha S_{i,j} + (1 - \alpha) S_{i,j-1} \quad \text{其中 } 0 \leq \alpha \leq 1$$

α 表示权重，最大值为 1，越大则预测越依赖刚过去 T 秒的流量。

4. 根据权利要求 2 或 3 所述的实现聚合链路流量均衡的方法，其特征在于，可以选取 $T=5$ 秒， α 为 0.8。

一种实现聚合链路流量均衡的方法

技术领域

本发明涉及网络设备之间多条物理链路捆绑成一条逻辑链路的聚合链路流量均衡的方法。

背景技术

交换机与交换机，交换机和服务器之间可以使用两条以上的物理链路进行连接。这样做的好处是增加了带宽，链路聚合的带宽等于聚合的物理链路带宽的总和。高可靠性，当其中某条物理链路发生故障时，交换机自动识别失效链路并停止从这条链路传送信息，同时选择其他链路继续进行传送。主干中一条或多条物理链路失效，不会影响两台设备的连通性，只是链路带宽随着失效链路数的增加而下降，因此链路聚合技术可以提高主干的可靠性。负载均衡，链路聚合技术有一个帧分发算法，可以根据各物理链路的流量有选择地进行数据包发送，使各链路的负载均衡。物理连接变化时可以快速配置或重配置。

交换机一般有多个端口，其主要功能是将一个端口接收的数据包转发到另外一个适当的端口中去。实现数据包的交换的基本原理是这样的：交换机中存储有叫做地址表的数据表格和一个学习模块。地址表至少包括一个地址项和一个端口项，表明为某一目的地址的数据包应该从对应的端口发送。交换机从一个端口接收到一个单播数据包后，首先提取出目的地址信息，在内部地址表中查询，如果能够找到对应的表项，则按照对应的端口转发该单播包，否则按广播包处理。与此同时学习模块会根据接收包的源地址和源端口学习并更新地址表项。

在美国专利 6049528 描述了实现链路聚合的方案。在设备中有一个地址表，在接受到一个数据包后，首先根据目的地址查找地址表，如果地址表中存在相应的表项，则按照地址表中的对应端口转发该数据包。如果没有对应的表项，则按照广播包处理。地址表项则是按照一般交换机的规则学习得到。美国专利申请号为 20020012340 和 20020010791 说明中描述了类似的系统，其交换转发

的规则是根据接收到的数据包的部分信息，即目的地址的某几位来决定从哪个端口转发。

以上所述的方法都存在一个问题，即地址学习时都是按照地址的几个比特来决定将该地址与哪个端口相联系。这种方法就有可能使得某些端口流量较大，而另外的端口流量很小，产生流量在各个端口之间分布不均衡的现象。

发明内容

本发明的解决的技术问题是提供了一种实现聚合链路流量均衡的方法，克服了单纯根据接收数据包的目的地址来决定地址表项的做法中存在的链路聚合中可能出现流量不平均的现象。

本发明所述的实现聚合链路流量均衡的方法，包括以下处理步骤：

- 步骤一，接收一个单播数据包，如果地址表中有相匹配的表项则转发；
- 步骤二，如果地址表中没有相匹配的地址项，则送往地址学习模块；
- 步骤三，地址学习模块根据端口的历史流量信息预测各端口未来的流量；
- 步骤四，选择未来流量最小的端口与该地址联系并更新地址表。

所述步骤三中预测端口未来流量具体包括以下处理步骤：

- a) 采用定时查询的方法，每隔 T 秒读取计数器中累计的端口发送字节数，假设端口 i 第 j 次查询的结果是 $C_{i,j}$ ；
- b) 计算端口 i 第 j 次查询前 T 秒的平均速率，计算公式为：

$$S_{i,j} = \frac{C_{i,j} - C_{i,j-1}}{T}$$

- c) 采用过去几个时间段流量的加权平均值来估计未来的流量。

如果采用两个时间段的流量预测未来 T 秒的平均速率，预测公式为：

$$S'_{i,j} = \alpha S_{i,j} + (1 - \alpha) S_{i,j-1} \quad \text{其中 } 0 \leq \alpha \leq 1$$

α 表示权重，最大值为 1，越大则预测越依赖刚过去 T 秒的流量。

一般情况下可以选取 $T=5$ 秒， α 为 0.8。

采用本发明的实现聚合链路流量均衡的方法，可以克服以往根据数据包地址的某几位来决定转发端口可能使不同端口负载不均衡的缺点，使转发数据包时可以更均衡地在物理链路上分配流量，更充分地利用物理链路提供的带宽。本方法还具有简单，速度快的优点。

附图说明

图 1 是两个系统（比如两个交换机之间或交换机与服务器之间）通过多条物理链路连接的情况。

图 2 是交换机或服务器的一个端口具有的功能模块。

图 3 是端口处理接收数据包的流程。

图 4 是本发明实现聚合链路流量均衡的方法的地址学习流程。

具体实施方式

下面结合附图来说明本发明的基本原理。

图 1 是两个系统（比如两个交换机之间或交换机与服务器之间）通过多条物理链路连接的情况。系统 A 的端口 1 至端口 4 与系统 B 的端口 4 至端口 1 相连，这 4 条链路聚合为一条逻辑链路。在本图中没有画出非聚合端口。

图 2 中画出了一个端口的功能模块。端口接收模块负责从端口上接收数据包。地址学习模块根据接收到数据包的源地址和接收端口号进行地址学习并更新地址表。地址表则记录了地址学习的结果，表明以某一地址为目的地址的数据包应该从哪个端口转发出去。在未设置 TRUNKING（中继）的情况下，CPU 软件处理模块不参与地址学习过程。包转发模块根据接收数据包的目的地址查询地址表并做出转发决策。端口发送模块实现数据包的发送。

图 3 是聚合端口处理接收数据包的流程。一个端口在设置成为聚合端口后，地址学习模块不再起作用。任何接收到的单播数据包先送到包转发模块，包转发模块根据包的目的地址在地址表中查询，如果在地址表中查到了匹配的表项，则进行正常的数据包转发。如果地址表中没有匹配的表项，则进行广播处理，并将地址消息发送到 CPU 软件处理模块进行地址学习。软件地址学习模块用学

习的结果来更新地址表。

图 4 是本发明实现聚合链路流量均衡的方法的地址学习流程。为了实现流量均衡的目的，首先预测各个端口未来的流量，并根据预测的流量决定学习的结果。CPU 收到地址学习消息，该地址消息包括接收到的数据包의 源地址和接收端口号等信息。查询聚合组中所有可用端口的流量，并预测未来端口的流量，选择流量最小的一个端口作为以该地址为目的地址的数据包的转发端口，并对地址表进行更新。

下面再详细介绍预测未来端口流量的实现方法：

交换机中都具有端口流量统计的能力，交换机的硬件提供一些 32 位的循环计数器，软件通过读计数器，可以知道端口累计发送了多少字节的数据。可采用定时查询的方法，每隔 T 秒中查询一下计数器，假设端口 i 第 j 次查询的结果是 $C_{i,j}$ 。

端口 i 第 j 次查询前 T 秒的平均速率为：

$$S_{i,j} = \frac{C_{i,j} - C_{i,j-1}}{T}$$

通常一个端口未来的流量同过去的流量有较大的相关性，所以我们可以采用过去几个时间段的流量的加权平均值来估计未来的流量。在下面的公式中，我们用过去两个时间段的流量预测未来 T 秒的平均速率：

$$S'_{i,j} = \alpha S_{i,j} + (1 - \alpha) S_{i,j-1} \quad \text{其中 } 0 \leq \alpha \leq 1$$

α 表示权重，最大值为 1，越大则预测越依赖刚过去 T 秒的流量。在我们对当前端口的未来流量作出预测之后，每次出现一个新地址需要学习时，我们不是根据地址本身来更新地址表项，而是将这个新地址与流量最小的端口联系起来，并更新地址表。在上述公式中，我们选取 T=5 秒， α 为 0.8 一般可以取得较好的效果。

以上详细介绍了采用本发明的方法在两个设备之间多条物理链路聚合成一条逻辑链路并实现负载均衡的方法。

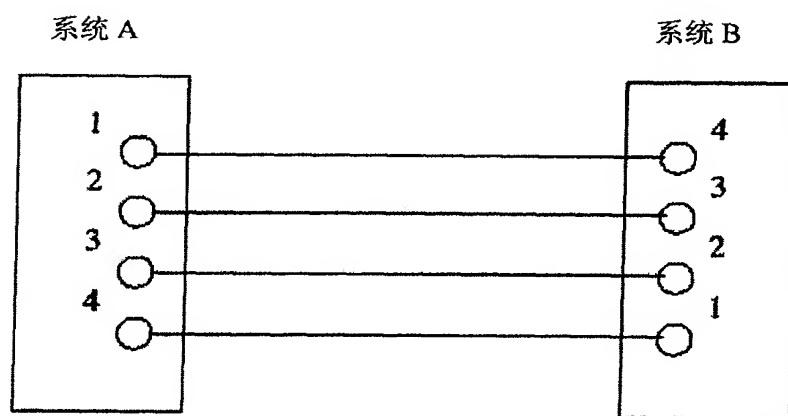


图 1

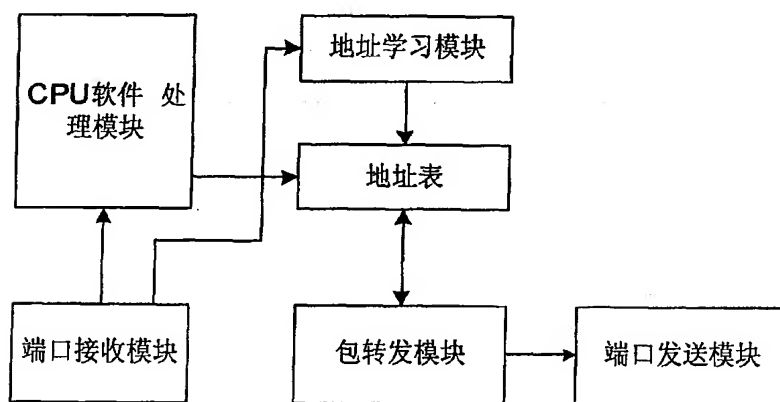


图 2

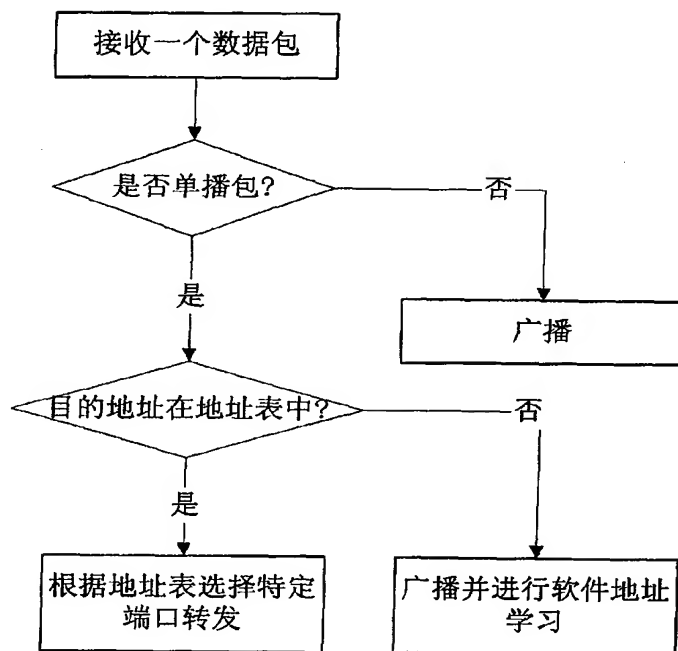


图 3

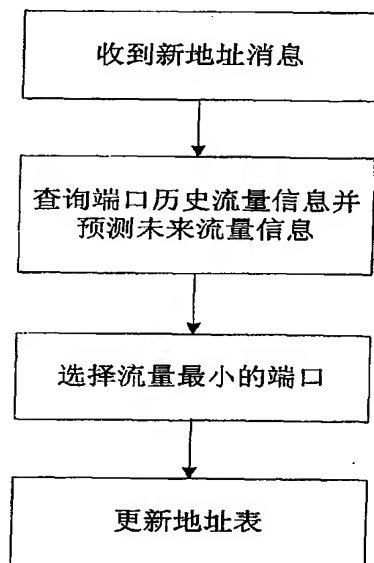


图 4